arXiv:nlin/0403006v1 [nlin.AO] 5 Mar 2004

Self Organized Scale-Free Networks from Merging and Regeneration

Beom Jun Kim,[1] Ala Trusina,[2,3] Petter Minnhagen,[3,2] and Kim Sneppen[3]

[1] *Department of Molecular Science and Technology, Ajou University, Suwon 442-749, Korea*
[2] *Department of Physics, Umeå University, 90187 Umeå, Sweden*
[3] *NORDITA, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark**

(Dated: February 5, 2008)

We consider the self organizing process of merging and regeneration of vertices in complex networks and demonstrate that a scale-free degree distribution emerges in a steady state of such a dynamics. The merging of neighbor vertices in a network may be viewed as an optimization of efficiency by minimizing redundancy. It is also a mechanism to shorten the distance and thus decrease signaling times between vertices in a complex network. Thus the merging process will in particular be relevant for networks where these issues related to global signaling are of concern.

PACS numbers: 89.75.-k

The ubiquitous broad degree distribution of the real world networks has been a matter of discussions for quite some time (see Ref.[1]-[9]). The question as to why broad degree distributions are observed in so many different networks, has triggered various proposals for their dynamical evolution. Roughly these proposals can be classified into two main scenarios: One is related to Simon's model of human behavior, Ref. [10], and was introduced in a network version under the name "preferential attachment" (see Ref. [11]). A related scenario is found in the protein duplication model Ref. [12] which is able to generate broad degree distributions because duplication, to some extent, mimics preferential attachment to neighboring nodes. Another class of models is where a scale-free distribution appears as a result of a balance between a modeled tendency to form hubs against an entropic pressure towards a random network with an exponential degree distribution. This approach includes direct attempts to construct Hamiltonians (see Ref. [13], [14]), local optimization approaches [15] as well as generation of scale-free networks by balancing a threshold for assigning links weighted according to exponentially distributed binding strengths [16].

In this paper we are presenting a new way of obtaining the scale-free degree distributions $P(k) \sim k^{-\gamma}$. The proposed mechanism is associated to the phenomena of aggregation with injection suggested in the context of astrophysical systems [17]. The model describes an evolving network, in which the main components, represented by nodes, are capable of pairwise merging, while at the same time the size of the network is maintained by generation of new nodes.

In real world networks one may think of the corresponding redistribution of links as a synergetic process associated with an increased efficiency in the linking pattern. For example, consider the network of interconnected computers. Since the computational power of the computers improves tremendously fast, periodically it could become more favorable to replace two out-dated neighboring server machines with one new machine that

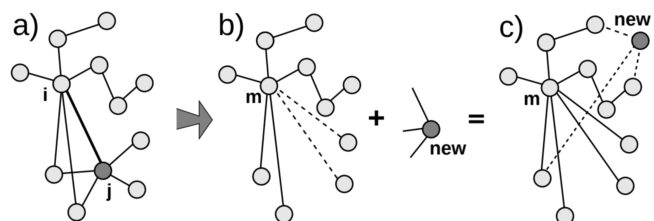

FIG. 1: A schematic representation of the update rule. (a) A node i is chosen at random and one of its neighbors j is randomly picked (k_i and k_j are the degrees of i and j, respectively). (b) Node m is a result of the pairwise merging with degree $k_m = (k_i - 1) + (k_j - 1) + N_{\mathrm{common}}$, where N_{common} is the number of common neighbors of i and j and the subtracted 1 is due to the lost common link. Node *new* is added with degree k_{new} from a uniform distribution and it attaches links to k_{new} random nodes (c).

can handle more connections. This simplifies the local network topology since the connections between the two old servers and the redundant links to other nodes are no longer needed. At the same time new servers may be constantly created to fulfill new demands. The generic merging or take-over process is defined by the update rule:

- At each step we choose the node i with degree k_i randomly, and then chose one of its random neighbors j. (See Fig.1a).

- The nodes i and j are merged together and thus a node m of degree $k_m = (k_i - 1) + (k_j - 1) - N_{common}$ appears instead, with N_{common} being the number of nodes that are neighbors to both i and j.

- At the same time a new node of some degree k_{new} is added to the network (Fig. 1b) with the links attached to k_{new} random nodes (see Fig.1c). The degree k_{new} of a newly added node is a random number r picked from a uniform distribution with average $\langle r \rangle$.

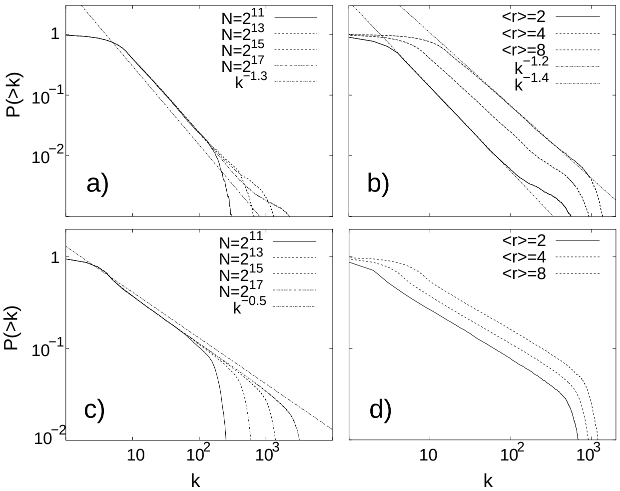

FIG. 2: (a) The cumulative node degree distribution $P(>k)$ for networks of sizes $N = 2^{11}$, 2^{13}, 2^{15} and 2^{17} with the average $\langle r \rangle = 8$. The fit is the power-law form $P(>k) \sim k^{-\gamma+1}$ with $\gamma = 2.3$. (b) $P(>k)$ for $\langle r \rangle = 2, 4$ and 8 and system size $N = 2^{14}$. The straight lines are the power-law fit with $\gamma = 2.4$ for $\langle r \rangle = 2$ and $\gamma = 2.2$ for $\langle r \rangle = 8$. (c) The cumulative degree distribution $P(>k)$ for four different system sizes for the realization when two randomly selected nodes are merging. The fit is $P(>k) \sim k^{-\gamma+1}$ with $\gamma = 1.5$. (d) $P(>k)$ for the merging of randomly selected nodes for $\langle r \rangle = 2, 4,$ and 8 and system size $N = 2^{14}$. The slope for every $\langle r \rangle$ is $\gamma = 1.5$.

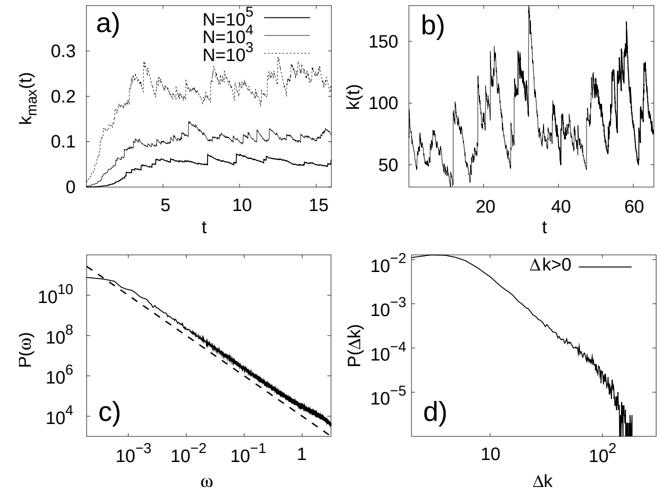

FIG. 3: Dynamics of the model network, with time measured as the number of updates per node. (a) Transient evolution of the maximal degree k_{max} in the system for different system sizes. Notice that the y-axis is normalized with system size N. (b) Degree $k(t)$ of a given random node as a function of time t. When the node is merged, $k(t)$ shows a sudden abrupt increase. (c) The power spectrum $P(\omega)$ versus the angular frequency ω obtained from the Fourier transformation of $k(t)$. The fit is a power-law with exponent -2. (d) The distribution of changes Δk for the involved nodes in each update.

Effectively this update reads:

$$\begin{cases} k_i \to k_i + k_j - N_{\text{common}} - 2 \\ k_j \to r \end{cases}, \qquad (1)$$

where in addition the N_{common} common neighbors are loosing one connection each, and r random nodes get one connection each. After the merging, i and j lose their identities and thus Eq. (1) can equally be written with i and j exchanged.

In Fig. 2(a) we show the cumulative degree distributions $P(>k)$ resulting from the update rule (1), which is the probability of finding a vertex with degree larger than k, for networks of different sizes at a steady state. The distribution is broad, and in fact clearly exhibits a broad range of power-law behavior from degree of about $k = \langle r \rangle$ up to a cutoff which increases with system size as shown in Fig. 2(a). The crucial point to note is that the scale-free network is an emergent property based on a simple merging process and that the driving mechanism is not related to preferential attachment.

In order to clarify this further we first note that the present neighbor-merging process (see Fig. 2) in some sense implicitly introduces a touch of "preferential" since, when taking a random neighbor of a random node, the neighbor is in some average sense selected with a probability proportional to its degree. However, this touch of "preferential" is not an essential part of why the merging generates scale-freeness as is illustrated by considering a version of the merging process where two random nodes are merged irrespective of whether or not they are connected, i.e. without any touch of "preferential". In that case one always, independently of the value of $\langle r \rangle$, obtains a scale-free distribution $P(>k) \propto 1/k^{0.5}$ (see Fig. 2c, d). Thus it is the merging, and not the preferential attachment that is the primary cause of the scale-free distribution. In fact it is remarkable that the neighbor-merging produces a narrower distribution than the completely random merging. (Compare Fig. 2b and Fig. 2b where $\gamma \sim 2.3$ for the neighbor-merging and $\gamma = 1.5$ for the random merging.) This reflects the property of merging to limit growth of hubs by their absorption of singly connected neighbors. (See the update rule (1), if $k_j = 1$ then $k_i \to k_i - 1$.) This tendency is stronger in the neighbor-merging process than in the random node merging due to the larger probability for a hub to merge with a single node in the former case. It means that the "touch of preferential" for the neighbor-merging actually inhibits the growth of a hub. This is in fact opposite to case of "preferential attachment" where hubs are thriving by accumulation of neighbors of low degree.

In the following we will discuss the original formulation of the mechanism (1). The main motivation being that the neighbor-merging version is likelier to be relevant for real networks since merging among the neighbors seems more natural than the merging of random nodes. In that case the only parameter in the system is the average de-

gree of the nodes, set by the average value of $\langle r \rangle$. In Fig. 2(b) we show $P(>k)$ for three different values of $\langle r \rangle$. The exponent γ in the power-law form $P(>k) \sim k^{-\gamma+1}$ decreases as one increases $\langle r \rangle$. For instance, $\gamma = 2.4$, and 2.2 for $\langle r \rangle = 2$ and 8 respectively. Furthermore we verified that in all cases the steady state degree distribution depends neither on the initial average degree nor on the shape of the initial degree distribution, be it a narrow distribution (star-like or exponential) or a broad one (scale-free).

An additional noteworthy feature of the model is that it produces networks without any degree-degree correlations. We measure the correlation profile $C(k_1, k_2) \equiv N(k_1, k_2)/N(k_1, k_2)_{\text{randomised}} - 1$, where $N(k_1, k_2)$ is the number of edges connecting nodes with degrees k_1 and k_2, and $N(k_1, k_2)_{\text{randomised}}$ is the corresponding quantity measured in the randomized network by many steps of edge exchanges (see Ref. [19] for details). We always find $|C(k_1, k_2)|$ to be small, $|C| < O(10^{-1})$, implying absence of significant degree-degree correlation in the networks emerging from the update rule (1).

The emergence of scaling is associated with a transient during which larger hubs are slowly forming, resulting in a self sustaining ecology with a broad degree distribution. This transient is illustrated in Fig. 3(a), where we follow the degree of the, at any time, most connected node in the system. This allows us to follow the transient approach towards the steady state. By data collapse (not shown) we found that the transient time increases slightly with system size, $\propto N^{0.2}$, whereas the maximum connected node at steady state has a degree, $k_{max} \propto N^{0.3}$. In Fig. 3(b) we follow a single node in steady state for a $N = 10^3$, and observe an intermittent behavior, which as seen in Fig. 3(c), can be characterized by a $1/\omega^2$ power spectrum. The power-law decay form of the power spectrum indicates the absence of a characteristic time scale, which is in parallel with the absence of the characteristic degree scale in the limit of large N observed in Fig. 2. We stress, that although the $1/\omega^2$ spectra resembles the one obtained for a random walk process, the actual dynamics is richer. This is reflected by the large jumps in increases of degree $k(t)$ in Fig. 3(b). This is quantified further by the broad distribution of changes $P(\Delta k)$ in Fig. 3(d).

So far we have been discussing a non-growing version of the network with the number of nodes being constant at each time step. One might argue that the majority of the real world networks are not in a steady state, but increase in size. For example, both the World Wide Web and the Internet are growing. Our merging algorithm can be extended to include a growth process if we add new nodes at each time at rate higher than that of merging. We stress that the growth is *non-preferential* in the sense that the newly added nodes link to the existing nodes with a probability that is independent of their degree. We start from a small initial network and grow it with $\langle r \rangle = 4$ at various values of the growth rate g until the network sizes reach $N = 10^5$. In Fig. 4 we show $P(>k)$ at the growth rates 0.1, 1.0, and 10. For example, if the growth rate is 0.1, one vertex is added per every 10 steps on average. If the network size increases very slowly, say one node per hundred basic steps, then the degree distribution approaches the one obtained for the non-growing case. As the growth rate is increased the distribution still retains its power-law form shape, but the slope γ increases to, e.g., 2.8 for the growth rate 1. As the growth rate is further increased, γ reaches 3, and then the power-law form begins to break down, and the degree distribution turns into the exponential one. The change in the slope reflects the difference in merging frequency and the frequency at which new nodes (typically nodes of low degree $k \sim \langle k \rangle$) are added to the system. In other words, the competition between the two time scales, one related with the merging and the other related with the growth, results in different degree distributions as the growth rate is changed. The overall feature is that the degree distribution becomes narrower at a higher growth rate because there is not enough time for the merging of the newly added nodes to spread across the whole network before the system grows further.

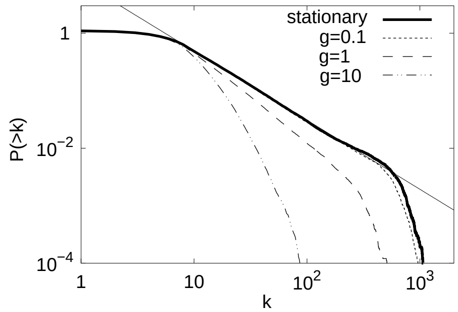

FIG. 4: The cumulative degree distribution for the growing network. Different curves correspond to different rates of growth. The solid curve is the degree distribution in non-growing case, and the line fit has slope $\gamma = -2.2$. With increasing the growth rate the distribution deviates from the stationary distribution: For moderate growth rates ($g = 0.1$ and 1) the distribution remains scale-free, whereas it collapses to exponential for larger growth rates ($g = 10$).

We also note that the fact that the merging and the regeneration mechanism gives rise to scale-free distributions does not hinge on the network structure per se. It is also applicable to entities characterized by just a scalar number, as is further discussed in [20].

In this Letter we propose a generic and robust mechanism for obtaining a broad, scale-free, degree distribution in networks where merging of nodes play a major role. The mechanism differs fundamentally from the preferential attachment mechanism [2] where a broad distributions are generated during gradual growth of hubs. The broad distribution resulting from merging and regeneration process emerges after a transient with slow building

up of a zoo of nodes of various degrees which, as the steady state is approached, together build up a scale-free distribution. We suggest that the mechanism could be relevant in a number of real world networks where the redistribution of links is associated with increasing efficiency in the linking pattern through minimization of pathway lengths. As an example we suggest that merging may be the effective result of evolution of architecture of protein regulatory networks in a cell. In these transcription and signalling networks, the time it takes to transmit signals is important [21], and it may thus be advantageous to eliminate an intermediate regulatory protein and move its regulation to an upstream regulatory protein. With the addition of new functions in form of new proteins (nodes), this effectively corresponds to the merging and regeneration model proposed in this paper.

A.T., P.M. and K.S. acknowledges the support of Swedish Research Council through Grants No. 621 2002 4135 and 629 2002 6258. B.J.K acknowledges the support by Korea Science and Engineering Foundation through Grant No. R14-2002-062-01000-0.

* Electronic address: trusina@tp.umu.se

[1] M. Faloutsus, P. Faloutsus, and C. Faloutsus, Comput. Commun. Rev. **29**, 251 (1999).
[2] R. Albert, H. Jeong, and A. -L. Barabási, Nature (London), **401**, 130 (1999).
[3] A. Broder *et al.*, Computer Networks **33**, 309 (2000).
[4] R. Pastor-Satoras, A. Vazques, and A. Vespegnani, cond-mat/0105161 (2001).
[5] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[6] M. E. J. Newman, Proc. Nat. Acad. Sci. **98**, 404 (2001).
[7] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[8] R. Ferrer, C. Janssen, and R. Sole, Phys. Rev. E **63**, 32767 (2001).
[9] R.V. Sole and J. M. Montoya, Proc. Roy. Soc. L. B **268**, 2039 (2001).
[10] H. A. Simon, Biometrica **42**, 425 (1955).
S. Bornholdt and K.H. Ebel, Phys. Rev. E. **64** R035104 (2001).
[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[12] A. Vazquez, A. Flammini, A. Maritan, and A. Vespegnani, cond-mat/0108043; R. V. Sole, R. Pastor-Satorras, E. D. Smith, and T. Kepler, Advances in Complex Systems **5**, 43 (2002).
[13] J. Berg and M. Lässig Phys. Rev. Lett. **89**(22),228701 (2002)
[14] M. Baiesi and S. S. Manna cond-mat/0305054
[15] M. Rosvall and K. Sneppen, Phys. Rev. Lett. (in press).
[16] A. Capocci, G. Caldarelli, R. Marchetti, and L. Pietronero Phys. Rev. E. **64** R035105 (2001);
[17] G.B. Field and W.C. Saslow, Astrophys. J. **142**, 568 (1965)
[18] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[19] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[20] P. Minnhagen, M. Rosvall, K. Sneppen and A. Trusina, preprint (2004).
[21] S.S. Shen-Orr, R. Milo, S. Mangan and U. Alon (2002). Nature Genetics, Published online: 22 April 2002, DOI:10. 1038/ng881